DOCUMENT RESUME

ED 050 147 TM 000 535

AUTHOR	Edwards, Keith J.
TITLE	Correcting Partial, Multiple, and Canonical Correlations for Attenuation.
INSTITUTION	Johns Hopkins Univ., Baltimore, Md. Center for the Study of Social Organization of Schools.
PUB DATE	Feb 71
NOTE	15p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Analysis of Variance, *Correlation, Error Patterns, Predictive Validity, *Predictor Variables, Reliability, *Research Tools, *Standard Error of Measurement, *Statistical Analysis, True Scores

ABSTRACT

The correction for attenuation formulas for partial, multiple, and canonical correlation coefficients are discussed and the effects of measurement errors on these statistics are explored. The notation is standardized and the derivation extended where appropriate. It is shown that as the reliabilities of the predictors become more disparate, the true contributions of each variable become more distorted. Relevant supporting formulas are included. (AE)

CORRECTING PARTIAL, MULTIPLE, AND CANONICAL CORRELATIONS FOR ATTENUATION

Keith J. Edwards
Center for Social Organization of Schools
The Johns Hopkins University
Baltimore, Maryland

A paper presented at the annual meeting of
The American Educational Research Association
New York, New York
February 4-7, 1971

CORRECTING PARTIAL, MULTIPLE, AND CANONICAL CORRELATIONS FOR ATTENUATION

Introduction

In dealing with multivariate correlational techniques and fallible data, one is faced with the same difficulties that have been pointed out for the product-moment correlation (Finucci, 1970) and other related measures of association (Stanley and Livingston, 1970). Correlations based on fallible variables will result in values which are underestimates of the correlations among the true parts of the variables. (The truth of this statement for all multivariate situations has not been proven analytically, but Cochran's (1970) work regarding multiple correlation suggests that the statement does, in fact, hold true.) Investigators have been inclined to ignore the problems of unreliability, being content with fallible underestimates of the true relationship and avoiding "questionable" correction for attenuation procedures. However, such an approach ignores useful information. In addition to providing a means for obtaining estimates of true score correlations, correction for attenuation formulas facilitate understanding of the effects of unreliability on the results. Information of this type is useful, for example, in deciding how much could be gained by expending time and money to develop more reliable measurement.

Size of the correlation coefficient is not the only concern for research involving multivariate measurement. One is often more concerned with the contribution of each individual variable to the overall result. For example, a canonical correlation analysis rarely stops with the coefficient itself. The relative sizes of the weighting factors for each of the variables in the canonical variates are vital for interpretation. Errors of measurement attenuate these weighting factors as well as the overall correlation, making the interpretation of canonical correlations and variates computed from fallible data a questionable, or at best difficult, undertaking. This same point holds true for multiple correlation and the use to which it is put.

The purpose of the present paper is to give the correction for attenuation formulae for partial, multiple, and canonical correlation coefficients and to discuss, where known, the effects of measurement error on these statistics. Most of the formulas presented have been derived elsewhere in the literature. I have simply standardized the notation and extended some of the derivation where appropriate.

The Partial Correlation Corrected for Attenuation

First, let us consider the first-order partial correlation coefficient. Suppose we have three variables x_1, x_2, and x_3 which are fallible measures of, say, alienation, school achievement, and I.Q. and want to know the true correlation between alienation and school achievement, controlling for I.Q. We begin by defining the variable x_i to be the sum of its true score, t_i, and errors of measurement, e_i (see (1) on list of formulae). We assume that $\sigma^2_{t_i e_i} = 0$, $\sigma^2_{t_i e_i} = 0$, and $\sigma^2_{e_i e_j} = 0$. That is, we begin with the classical test theory model and the classical test theory assumption.

The partial correlation between alienation and achievement, controlling for I.Q., is defined as the zero-order correlation of residuals. The residuals for alienation are given by the difference between the observed values and the regression estimates of alienation from I.Q. The residuals are represented symbolically in equations (2) and (3). It is well-known that the correlation of residuals can be expressed in terms of the three zero-order correlations. The formula is given by (4), which is the partial r based on fallible variates. The partial correlation coefficient, corrected for attenuation, would yield the partial correlation of true score, i.e., the correlation of true score residuals. We can obtain the correction for attuation formula by starting with the correlation of true score residuals and working backwards.

The true score residuals are defined as the difference between the true value and the estimated true value based on a regression of the variable (t_1 or t_2) on the true value of the control variable (t_3) and are given in formulas (5) and (6). The partial correlation of t_1 and t_2 controlling for t_3 is then given by (7). Expanding numerator and denominator, we can use some of the well-known properties of classical test theory to express the true partial correlation in terms of the fallible zero-order correlations and reliabilities. The result, given in (11) is the correction for attenuation formula for a first-order partial correlation. It should be noted that the formula is equivalent to correcting each of the zero-order correlations for attenuation by the usual way and plugging these values into (4). (See Livingston and Stanley, 1970.)

Bohrnstedt (1969) derived a formula for correcting partial correlations for attenuation due to errors of measurement which is similar to (11), but does not contain the terms ρ_{11}, ρ_{22}. Upon examining his derivation, it was apparent that he was correcting only for errors of measurement in the control variable x_3. In effect, he had provided the formula for a partially corrected partial correlation coefficient. On the basis of his formula, Borhnstedt indicates that it is possible for the corrected partial correlation to be less than the obtained partial correlation. This does not seem to be the case however, and it appears that correcting for attenuation will result in larger values. Since $\rho_{33} \leq 1$, the numerator of (11) will be less than or equal to that of (4). This would tend to decrease the value of the corrected partial correlation. On the other hand, the denominator of (11), being less than that of (4), would tend to increase $\rho_{T_1T_2 \cdot T_3}$. The relative size of the numerator to the denominator in (11) would seem to result in an overall increase in the partial correlation, when corrected for attenuation. (A few numbers I have plugged into the equations indicate such a trend, though I have no analytic proof of this statement.)

A General Approach to Multivariate Corrections for Attenuation

Meredith (1964) has developed a more general approach to correction for attenuation problems which he has applied to the canonical correlation problem. His result can be readily applied to problems involving partial and multiple correlation. We begin with a variance-covariance matrix, $\sum_x$, of rank p + q, where p + q is the number of variables being fallibly measured. Under the assumption that the classical test theory model is appropriate for each of the p + q variabes we can write the matrix $\sum_x$ as the sum of two matrices, $\sum_t$, the variance-covariance matrix among true scores, and $\sum_e$, the variance-covariance matrix among the errors of measurement (equation 12). Assuming errors of measurement to covary zero with each other the matrix $\sum_e$ is a diagonal matrix of the variance errors of estimates. We can obtain $\sum_t$ by subtraction (equation 13). Given $\sum_t$, the matrix of true score variances and covariances, it is a simple matter to obtain the matrix of true score correlations by dividing each element by the square root of the product of the appropriate variances. These operations are shown in matrix notation in equation (14).

It is important to note here that (14) is equivalent to correcting each of the zero-order correlations in P_x, the matrix of fallible correlations, for attenuation in the usual manner. That such is the case becomes clear if we consider each of the p + q variables to have mean=zero and variance=one. Under these conditions $\sum_x = P_x$ and $\sum_e$ is a diagonal matrix of alienation coefficients. Thus, the matrix $\sum_t$ of (13) is the matrix of fallible intercorrelations (P_x) with reliabilities on the diagonal, which is the true-score variance-covariance matrix of standard deviates. The operations shown in (14) now involve dividing every correlation in P_x by the square root of the product of the reliabilities for the appropriate variables, which is the zero-order correction for attenuation procedure.

So far, the discussion has been in terms of population values. Merredith has pointed out that a maximum likelihood estimate of $\sum_t$ and thus of P_t can be obtained from S_x, the sample variance-covariance matrix, if the reliabilities of the measures are known (equations 15 and 16). Though the remainder of the paper continues to use the population values, one can easily substitute $\hat{P}_t$ under the above restriction.

A general procedure for correcting multivariate correlations for attenuation involves the following two steps. First, correct each of the zero-order correlations for attenuation in the usual way to obtain P_t. Second, calculate the desired statistic from P_t.

Let us return to the problem of partial correlations. Suppose that we were interested in obtaining the true score correlations among a set of p variates controlling for true scores on a second set of q variables. We could solve the problem by first obtaining P_t (or, more likely, its estimates, $\hat{P}_t$), partitioning P_t as shown in (17), and using the matrix solution for partial correlations (Anderson, 1958, and Morrison, 1967) shown in (18). If $q = 1$, $P_{t_{1.2}}$ is a p x p matrix of first order partials whose off-diagonal elements are of the form given in (4). Extending our three-variabe example, $P_{t1.2}$ could be a matrix of attitude-achievement true score correlations, controlling for I.Q.

The multiple correlation problem involves finding the maximum correlations R between a single criterion and a linear combination of, say, p predictors. The matrix solution for R^2 is given in (19) (See Anderson, 1958 or Morrison, 1967). The multiple correlation between the true scores of the p predictors and the criterion could be obtained by substituting the corresponding true score correlation matrices of (17) into (19), resulting in equation (20).

In the above situation, any of the p + 1 variables could be designated as the criterion by simply interchanging the appropriate rows and columns of P_t. A general formula for the squared multiple correlation coefficient (SCM) of each of the i variates with the remaining q variates corrected for attenuation is given by (21), where I is a p + 1 identity matrix and D indicates diagonals of the matrices given in parentheses. A connection can be made here with factor analysis. It is common practice to factor a matrix of the form $[P - D(P^{-1})]$, which is the case where the SCM coefficient is used as an estimate of communality. (However, Harris (1964) indicates that such a procedure does not represent "true" factor analysis).

The last statistic we shall discuss is the canonical correlation coefficient (Hotelling, 1936). Canonical correlation is a generalization of the concept of multiple correlation to the case of multiple criteria ($q \geq 1$) as well as multiple predictors ($p \geq 1$). The objective in such an analysis is to find the maximum correlation between a linear composite of the predictors and a linear composite of the criteria. Though Hotelling was primarily concerned with the largest correlation between these composites, there are k = min (p, q) possible independent correlations. The k canonical correlations for any given set of p predictors and q criteria are given by the roots of the determinantal equation given in (22). If the true-score correlation matrix of (17) is used, you would have the k canonical correlations corrected for attenuation. For completeness, the formulas for the weighting vectors to form the linear composite of the criterion variates ($\underset{\sim}{a}_i$ in 24) and the linear composite of predictor variabes ($\underset{\sim}{b}_i$ in 25) are also given. The formulas would provide either the fallible weighting vectors or the true-score weights, depending on which correlation matrices were used.

Effects of Errors of Measurement

In the introduction it was pointed out that the most well-known effect of errors of measurement is to produce a statistic which is an underestimate of the true value. For example, Cochran (1970) has shown that for a number of situations involving multiple correlation, a good estimate of attenuating effects of fallible data is given by (22), though the actual value of R_x^2 may run 25% higher than this value for positive ρ_{iq} and low predictor reliability ($\rho_{ii} = 0.5$).

A second problem raised when errors of measurement are present is the valid interpretation of results for multivariate correlations. In multiple and canonical correlations studies an important objective is to discover the relative importance of the predictor and criterion variables. The intercorrelations among these variables and their unreliability can interact to produce misleading results. An example from Cochran (1970) illustrates this point.

A common practice in the application of multiple correlation (especially among sociologists) is to partition the predicted variance (R^2) into portions uniquely attributable to each predictor and the portion of common variance predicted. Unreliability can have a substantial effect on the results of such an analysis. Consider the 2-predictor case with $\rho_{13} = \rho_{23} = 0.5$, and $\rho_{12} = 0.$ and no error of measurement:

$R^2_{3.12} = .385$

% of variance unique to x_1 = 13.5
% of variance unique to x_2 = 13.5
% of variance common to both = 11.5

With the reliability of variable 1 equal to .8; i.e. $\rho_{11} = 0.8$ and $\rho_{22} = 1$,

$R^2_{3.12} = .356$

% of variance unique to x_1 = 10.6
% of variance unique to x_2 = 15.6
% of variance common to both = 9.4

With $\rho_{11} = 0.6$ and $\rho_{22} = 1$

$R^2_{3.12} = .328$

% of variance unique to x_1 = 7.8
% of variance unique to x_2 = 17.8
% of variance common to both = 7.2

In the above example we see that as the reliabilities of the predictors become more disparate, the true contributions of each variable becomes more distorted. This effect can be best understood when you consider what would happen if x_1 were removed from the correlation. $R^2_{3.2} = .25$ and percent of variance unique to x_2 would be 25. Unreliability in one of the variables takes part of that variable "out" of the prediction, shifting predicted variance to the more reliable predictors. The change in the importance of predictors in multiple correlation caused by deletion of one of the variables has been referred to as the "bouncing betas." Difference in the reliabilities adds more bounce to these results.

BIBLIOGRAPHY

Anderson, Theodore W. An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons Inc., 1958.

Bohrnstedt, George W. "Observations on the Measurement of Change" in Sociological Methodology, Edgar F. Borgatta (ed.). San Francisco: Jossey-Bass Inc., 1969. pp. 113-133.

Cochran, William G. "Some Effects of Errors of Measurement on Multiple Correlation" in Journal of the American Statistical Association, 65:22-34, March, 1970.

Finucci, Joan M. "Early History and Basic Derivation Formulas." A paper read at the American Educational Research Association Convention. New York: February 5, 1971.

Harris, Chester W. "Some Recent Developments in Factor Analysis" in Educational and Psychological Measurement, 26:193-206, 1964.

Hotelling, Harold. "Relations Between Two Sets of Variates," Biometrika 28:321-377, 1936.

Livingston, Damuel A. and Julian C. Stanley. "What is the True Coefficient of Partial Correlation?" Unpublished manuscript, 1970.

Meredith, William. "Canonical Correlations With Fallible Data, "Psychometrika, 29:55-65, March, 1964.

Morrison, Donald F. Multivariate Statistical Methods. New York: McGraw-Hill, 1967.

Stanley, Julian C. and Samuel A. Livingston. "Correcting Eta Squared, Epsilon Squared, Omega Squared, and Intraclass *r* for Attenuation" in paper read at the American Educational Research Association Convention, New York: February 5, 1971.

Formulas for Paper

Correcting Partial, Multiple, and Canonical Correlation for Attenuation

Keith Edwards

The Johns Hopkins University

(1) $x_i = t_i + e_i \qquad$ (for $i = 1, 2, 3$)

(2) $x_{1.3} = x_1 - \beta_{x_1x_3}x_3$

(3) $x_{2.3} = x_2 - \beta_{x_1x_3}x_3$

(4) $$\rho_{x_1x_2 \cdot x_3} = \frac{\sigma_{x_{1.3}x_{2.3}}}{\sqrt{\sigma^2_{x_{1.3}} \cdot \sigma^2_{x_{2.3}}}} = \frac{\rho_{x_1x_2} - \rho_{x_1x_3}\rho_{x_2x_3}}{\sqrt{(1 - \rho^2_{x_1x_3})(1 - \rho^2_{x_2x_3})}}$$

(5) $t_{1.3} = t_1 - \beta_{t_1t_3}t_3$

(6) $t_{2.3} = t_2 - \beta_{t_2t_3}t_3$

(7) $$\rho_{t_1t_2 \cdot t_3} = \frac{\sigma_{t_{1.3}t_{2.3}}}{\sqrt{\sigma^2_{t_{1.3}} \cdot \sigma^2_{t_{2.3}}}} = \frac{\sigma_{(t_1 - \beta_{t_1t_2}t_3)(t_2 - \beta_{t_2t_3}t_3)}}{\sqrt{\sigma^2_{(t_1 - \beta_{t_1t_3}t_3)}\sigma^2_{(t_2 - \beta_{t\ t_3}t_3)}}}$$

Derivation of correction formula for partial correlation coefficient.

Expanding the numerator of (7)

$$\sigma_{t_{1.3}t_{2.3}} = \sigma_{t_1t_2} - \beta_{t_1t_3}\sigma_{t_2t_3} - \beta_{t_2t_3}\sigma_{t_1t_3} + \beta_{t_1t_3}\beta_{t_2t_3}\sigma^2_{t_3}$$

$$= \sigma_{t_1}\sigma_{t_2}\rho_{t_1t_2} - \left(\rho_{t_1t_3}\frac{\sigma_{t_1}}{\sigma_{t_3}}\right)\sigma_{t_2}\sigma_{t_3}\rho_{t_2t_3} - \left(\rho_{t_2t_3}\frac{\sigma_{t_2}}{\sigma_{t_3}}\right)\sigma_{t_1}\sigma_{t_3}\rho_{t_1t_3}$$

(cont.)

---

*Presented at the American Educational Research Association 55th Annual Meeting, New York City, February 4-7, 1971.

$$+\left(\rho_{t_1t_3}\frac{\sigma_{t_1}}{\sigma_{t_3}}\right)\left(\rho_{t_2t_3}\frac{\sigma_{t_2}}{\sigma_{t_3}}\sigma^2_{t_3}\right)$$

(The last two terms are equivalent except for sign, and thus they sum to zero.)

$$= \sigma_{t_1}\sigma_{t_2}\rho_{t_1t_2} - \rho_{t_1t_3}\sigma_{t_1}\sigma_{t_2}\rho_{t_2t_3}$$

$$= \sqrt{\rho_{11}\rho_{22}}\sigma_{x_1}\sigma_{x_2}\frac{\rho_{x_1x_2}}{\sqrt{\rho_{11}\rho_{22}}} - \frac{\rho_{x_1x_3}}{\sqrt{\rho_{11}\rho_{33}}}\left(\sqrt{\rho_{11}\rho_{22}}\sigma_{x_1}\sigma_{x_2}\right)\frac{\rho_{x_2x_3}}{\sqrt{\rho_{22}\rho_{33}}} ,$$

Therefore, where ρ_{ii} is the reliability coefficient of variable x_i

$$(8) \quad \sigma_{t_{1.3}t_{2.3}} = \sigma_{x_1}\sigma_{x_2}\left(\frac{\rho_{33}\rho_{x_1x_2} - \rho_{x_1x_3}\rho_{x_2x_3}}{\rho_{33}}\right) .$$

Expanding the first term in the denominator of (7) one secures

$$\sigma^2_{t_{1.3}} = \sigma^2_{t_1} + \beta^2_{t_1t_3}\sigma^2_{t_3} - 2\beta_{t_1t_3}\sigma_{t_1t_3}$$

$$= \sigma^2_{t_1} + \rho^2_{t_1t_3}\sigma^2_{t_1} - 2\rho_{t_1t_3}\sigma^2_{t_1}\rho_{t_1t_3}$$

$$= \sigma^2_{t_1} - \rho^2_{t_1t_3}\sigma^2_{t_1}$$

$$= \rho_{11}\sigma^2_{x_1} - \frac{\rho^2_{x_1x_3}\rho_{11}\sigma^2_{x_1}}{\rho_{11}\rho_{33}} .$$

Therefore,

$$(9) \quad \sigma^2_{t_{1.3}} = \sigma^2_{x_1}\frac{\left(\rho_{11}\rho_{33} - \rho^2_{x_1x_3}\right)}{\rho_{33}} .$$

Similarly, it can be shown that

$$(10) \quad \sigma^2_{t_{2.3}} = \sigma^2_{x_2}\frac{\left(\rho_{22}\rho_{33} - \rho^2_{x_2x_3}\right)}{\rho_{33}} .$$

Substituting (8), (9), and (10) into (7) and simplifying we get

$$(11) \quad \rho_{t_1t_2.t_3} = \frac{\rho_{33}\rho_{x_1x_2} - \rho_{x_1x_3}\rho_{x_2x_3}}{\sqrt{\left(\rho_{11}\rho_{33} - \rho^2_{x_1x_2}\right)}\sqrt{\left(\rho_{22}\rho_{33} - \rho^2_{x_2x_3}\right)}} .$$

$$(12) \quad \Sigma_X = \Sigma_T + \Sigma_E$$

$$(13) \quad \Sigma_T = \Sigma_X - \Sigma_E$$

$$(14) \quad P_T = D(\Sigma_T)^{-1/2}\Sigma_T D(\Sigma_T)^{-1/2}$$

$$(15) \quad \hat{\Sigma}_T = S_X - \Sigma_E$$

$$(16) \quad \hat{P}_T = D(\hat{\Sigma}_T)^{-1/2}\hat{\Sigma}_T D(\hat{\Sigma}_T)^{-1/2}$$

$$(17) \quad P_T = \left[\begin{array}{c|c} P_{T_{11}} & P_{T_{12}} \\ \hline P_{T_{21}} & P_{T_{22}} \end{array}\right]$$

$$(18) \quad P_{T_{1.2}} = P_{T_{11}} - P_{T_{12}} P^{-1}_{T_{22}} P_{T_{21}}$$

$$(19) \quad R^2_X = P_{X_{21}} P^{-1}_{X_{11}} P_{X_{12}}$$

$$(20) \quad R^2_T = P_{T_{21}} P^{-1}_{T_{11}} P_{T_{12}}$$

$$(21) \quad D(R^2_{T_i}) = I - D(P^{-1}_T)$$

$$(22) \quad \left| P_{X_{21}} P^{-1}_{X_{11}} P_{X_{12}} - \lambda P_{X_{22}} \right| = 0$$

$$(23) \quad \left| P_{T_{21}} P^{-1}_{T_{11}} P_{T_{12}} - \lambda P_{T_{22}} \right| = 0$$

$$(24) \quad \left(P_{T_{21}} P^{-1}_{T_{11}} P_{T_{12}} - \lambda_i P_{T_{22}} \right) \underset{\sim}{a}_i = \underset{\sim}{0}$$

$$(25) \quad \underset{\sim}{b}_i = \frac{1}{\sqrt{\lambda_i}} P^{-1}_{T_{11}} P_{T_{12}} \underset{\sim}{a}_i$$

$$(26)\quad R_X^2 = R_T^2 \cdot \rho_{qq} \frac{\sum_{i=1}^{q-1} \rho_{iq}\rho_{ii}}{\sum_{i=1}^{q-1} \rho_{iq}} = R_T^2 \rho_{qq} \bar{\rho}_{ii}$$

where ρ_{qq} is reliability of the criterion

ρ_{ii} is reliability of the ith predictor

ρ_{iq} is correlation between criterion and ith predictor.